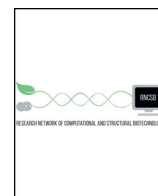




COMPUTATIONAL
AND STRUCTURAL
BIOTECHNOLOGY
JOURNAL

journal homepage: www.elsevier.com/locate/csbj



Mini Review

Blood transcriptomics and metabolomics for personalized medicine

Shuzhao Li ^{a,*}, Andrei Todor ^a, Ruiyan Luo ^b

^a Department of Medicine, Emory University School of Medicine, 615 Michael Street, Atlanta, GA 30322, USA

^b Division of Epidemiology and Biostatistics, School of Public Health, Georgia State University, One Park Place, Atlanta, GA 30303, USA

ARTICLE INFO

Article history:

Received 2 June 2015

Received in revised form 5 October 2015

Accepted 23 October 2015

Available online 31 October 2015

Keywords:

Transcriptomics

Metabolomics

Blood systems biology

Personalized medicine

Data integration

ABSTRACT

Molecular analysis of blood samples is pivotal to clinical diagnosis and has been intensively investigated since the rise of systems biology. Recent developments have opened new opportunities to utilize transcriptomics and metabolomics for personalized and precision medicine. Efforts from human immunology have infused into this area exquisite characterizations of subpopulations of blood cells. It is now possible to infer from blood transcriptomics, with fine accuracy, the contribution of immune activation and of cell subpopulations. In parallel, high-resolution mass spectrometry has brought revolutionary analytical capability, detecting >10,000 metabolites, together with environmental exposure, dietary intake, microbial activity, and pharmaceutical drugs. Thus, the re-examination of blood chemicals by metabolomics is in order. Transcriptomics and metabolomics can be integrated to provide a more comprehensive understanding of the human biological states. We will review these new data and methods and discuss how they can contribute to personalized medicine.

© 2015 Li et al.. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. An overdue review of “blood systems biology”	2
2. Data acquisition of transcriptomics and metabolomics	2
3. Untangling mixed cell populations in blood transcriptomics.	3
4. Metabolomics for disease markers	3
5. Pathways and modules—power in groups	3
6. Integrating different data types to understand disease pathophysiology	4
7. Concluding remarks.	5
Acknowledgment	5
References	5

Many human diseases are complex and heterogeneous, whereas diagnostic methods are still limiting. Genetics and high-throughput molecular profiling now helps to redefine the disease classifications [1,2]. Personalized and precision medicine aims to design therapeutic interventions based on the condition of individual patients. For example, in the case of *trastuzumab*, a drug that is administered to breast cancer patients, its therapeutic efficiency varies depending on the patient's breast cancer subtype. This is because *trastuzumab* targets HER2 (human epidermal growth factor receptor type 2) proteins, and it is only effective on breast cancers with HER2 overexpression [3]. Therefore, a diagnostic test that determines HER2 overexpression is required before

trastuzumab can be subscribed. A different type of example is adoptive T cell transfer for cancer immunotherapy, where specific T cells from an individual patient are engineered and expanded, then infused back to the same patient [4–6]. This type of therapy is “double” personalized because the T cells have to be from the very patient to be immunologically tolerant, and their surface receptors have to be specific to the tumor mutation found in that patient. Numerous examples exist that drug efficacy is limited due to the lack of “precision” mechanism. The widely used statins (cholesterol lowering drugs) may be efficacious in only 5% of the population, while esomeprazole (for heartburn treatment) fares even less [7]. A lot of research efforts have gone to identifying genetic variations associated with diseases, including many large genome-wide association studies (GWAS). However, genetic variations only account for small percentages of the occurrence of common

* Corresponding author.

E-mail address: shuzhao.li@gmail.com (S. Li).

diseases [8,9]. It is increasingly recognized that there is a large gap between genomics and phenotypes and that transcriptomics and metabolomics are important to fill this gap [10–14]. In this article, we will review the latest progress in transcriptomics and metabolomics, with a focus on samples from blood, a key tissue for clinical diagnosis. Since abundant introductory literature can be found on omics technologies and their data analysis, this article focuses more on important recent developments and opportunities.

1. An overdue review of “blood systems biology”

Blood has been intensively investigated since the beginning of molecular systems biology. Publications on disease diagnosis using blood transcriptomes are now numbered in thousands. Although it is widely recognized that mRNA only provides a slice of information from complex biology, few papers attempted to quantify the cell-level complexity in blood transcriptomics. Because blood is a mixture of many different cell types (Fig. 1), the fluctuation of cell populations alone causes large variations in transcriptomics data. This problem only became tractable with the recent progress in human immunology, where transcriptomics of isolated cell populations provided necessary information [15–17]. Nonetheless, a review on “blood systems biology” is long overdue.

As part of the body circulatory system, blood reflects the homeostasis of metabolism, hematopoietic development, and immune functions. As Fig. 1 shows, this involves many cell types and subtypes, and a number of “omics” technologies are employed to measure on different aspects of the system. The global molecular profiles of different cell types are tightly related to their developmental lineage and functions. As Novershtern et al. [18] showed, the clustering of transcriptomics data of blood cells reflects the hematopoietic process. The white blood cells are also sensitive indicators of the immune status. An infection will readily induce the influx of immune cells to blood as well as the activation of molecular programs in these cells. Cytokines and chemokines can increase dramatically during such events. The plasma contains molecular signals and wastes from the lymphatic system. The metabolites within plasma can reflect liver or kidney function, endocrine signaling, inflammation, and metabolic disorders. Thus, blood systems biology needs to address the following: (1) mixture data—most commonly, omics data are collected on peripheral blood mononuclear cells, where cell population composition is critical; (2) connection to a systemic model, such as pharmacokinetics or host–pathogen interaction models—blood is not a closed system by itself, only a window to systemic

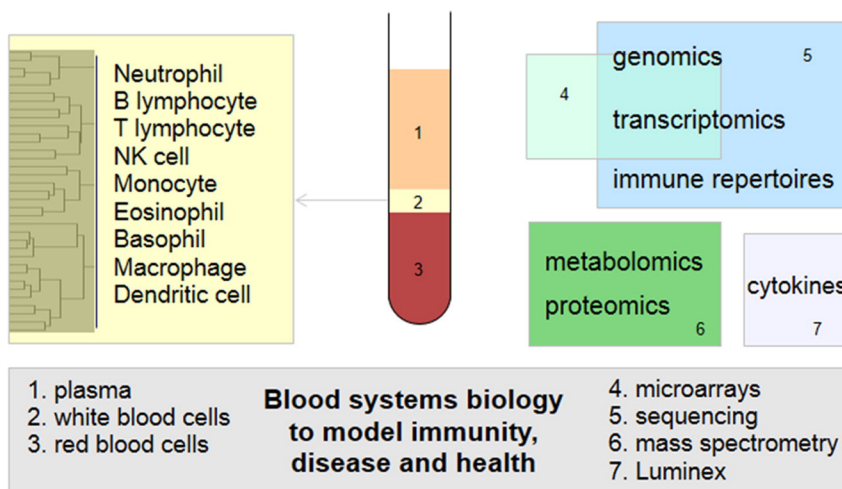
events; and (3) data integration. This could be the association between omics data and phenotype or the connection between different omics data types. We will start with an overview of transcriptomics and metabolomics then move on to specific topics for “blood systems biology”.

2. Data acquisition of transcriptomics and metabolomics

DNA microarrays were developed in the 1990s as a major technology to measure transcriptomics. The technology relies on the specific hybridization between complementary polynucleotides. Probes are designed based on known gene transcripts and tethered on a glass surface. Targets are generated from biological samples, labeled directly or indirectly with fluorescent dyes. The hybridization reactions are carried on in miniaturized chambers. After the probes capture specific targets, the fluorescent signals are scanned and reported based on their grid locations. Thousands of microarray experiments are now deposited in public repositories such as GEO [19] and ArrayExpress [20].

As the cost of DNA sequencing drops, RNAseq becomes a viable alternative to capture transcriptomics. Using massively parallel sequencing platforms, RNAseq reads the number of DNA copies that are converted from mRNA, thus quantifying the concentration of mRNA species. From these sequencing reactions, the sequence variations in exons, such as single nucleotide polymorphisms (SNPs) and alternative splicing, are also captured in the data. Both the experimental methods and the computational analysis of RNAseq are evolving rapidly, and significant improvements are expected.

Metabolomics is the global profiling of small molecules (usually under 2000 Da). While nuclear magnetic resonance (NMR) [21] has been a powerful tool, mass spectrometry coupled with liquid or gas chromatography is the most popular platform due to the superior sensitivity and coverage [22–24]. The newest high-resolution mass spectrometer, in particular, yields unparalleled precision in analyzing chemicals in complex biological samples. The basic principle used by mass spectrometers is the differentiated deflection of charged particles in a magnetic field based on their mass. By the Lorentz law, the magnitude of the deflection is proportional to the mass to charge ratio. The advanced version, Fourier transform mass spectrometers, can achieve spectacular mass resolution by measuring the spinning frequency of ions that are trapped and oscillate in a chamber. The computational aspects of metabolomics are also in rapid progress, including open source feature extraction tools (XCMS [25], OpenMS [26], apLCMS [27], xMSanalyzer [28]), databases of metabolites (Human Metabolome



Database [29,30], METLIN [31], PubChem [32], ChEBI [33]), and data analysis tools (XCMS Online [34], MetaboAnalyst [35], mummichog [36]). It should be noted that these data contain more than endogenous metabolites, also including food intake, microbial activities, pharmaceutical drugs, and environmental exposures. The collective measurement is sometimes termed as “exposome” [37,38].

For the analysis of both transcriptomics and metabolomics data, general principles of “omics” apply. With measurement of thousands of features, multiple test correction is necessary to control false positive rates [39,40]. The number of features is usually far larger than the number of samples. Statistical methods often “borrow” information from variation between features to help the estimation of variation between samples [41]. Prior knowledge of molecular pathways and interactions can be of great value, and the methods usually involve over-representation tests or network modeling [42,43]. Since these areas can be referred to other more generic reviews, we will highlight a few areas that are more pertinent to blood data: how to deal with data from the mixture of blood cells, emerging metabolomics data on plasma or serum, and useful pathway and network tools.

3. Untangling mixed cell populations in blood transcriptomics

When transcriptomics data are measured on a mixture of multiple cell populations, it is a reasonable assumption that the data are a linear combination of transcriptomes of each populations [44]. These separate cell populations can be obtained by flow cytometry-based sorting, and large quantity of data are made available in ImmGen and ImmPort databases [15,16]. Conversely, if the percentages of each cell population are known, variations may be attributed to each population by regression methods [45].

Since “omics” data are often noisy, pre-filtered cell-type-specific genes (markers) are very useful in this context [46–48]. The use of too few markers, like those in flow cytometry, is not recommended in transcriptomics analysis because a larger number of genes are needed to counter the measurement noise, and time differential may exist between protein (used in flow cytometry) and mRNA (measured in transcriptomics) levels. A set of cell-type-specific genes are included in the blood transcription modules from Li et al. [46]. With cell-type-specific markers, a statistical test of over-representation can reveal what cell type contributes to the most differential genes [42,48]. An example is shown in Fig. 2A: immunization using MCV4 vaccine upregulated 466 significant genes after 1 week. These genes contain 7 out of 24 signature genes for plasma cells, the major antibody secreting cells. Given that these numbers were drawn from genome-wide measurement of 20,722 genes, the enrichment on plasma cell signature genes is highly significant ($p < 10^{-5}$, Fisher exact test). Alternatively, one can leverage the GSEA (Gene set enrichment analysis [49]) statistical framework, using cell-specific markers as gene sets. This method shows that the same plasma signature of 24 genes are highly enriched for upregulated genes (p -value approaching 0, Fig. 2B). The GSEA approach can be more sensitive than over-representation tests and less biased by cutoffs in feature selection. In general, we have found that distribution tests in the style of Kolmogorov–Smirnov test suit well for assigning cell type information from blood transcriptomics, and the results are very consistent with flow cytometry data that were obtained on the same samples (unpublished).

4. Metabolomics for disease markers

While transcriptomics analysis usually requires cell collection protocols in place to preserve the integrity of mRNA, metabolomics is amenable to most archival samples. This easy access to samples and the reasoning that metabolites provide functional readout of gene activities gather a great deal of enthusiasm to look for disease markers using metabolomics [51–65]. Examples of metabolomics for biomarker study include diabetes [62,66], macular degeneration [67], asthma

[68], Parkinson's disease [69], nonalcoholic fatty liver disease [70], and tuberculosis [71]. Notably, metabolite markers of diabetes were reported many years prior to the disease onset [61]. The field of high-resolution metabolomics is advancing very rapidly [24,72]. Although it has been difficult to compare earlier data from different platforms, the accumulation of high-resolution metabolomics data may be approaching a critical threshold of assembling a reference human metabolome.

The current clinical blood tests report a limited number of metabolites (Fig. 3), most of which are detected in current metabolomics data. That is, with similar cost, metabolomics can already deliver quantitative information on hundreds of metabolites. The normal and abnormal ranges of many metabolites are either already in the literature or can be learned from large cohorts. Recently, Miller et al. [73] have already demonstrated that a single metabolomic analysis successfully diagnosed 20 inborn metabolic diseases. The potential of clinical metabolomics is revolutionary—once proofs of new disease markers sink in and regulatory approval comes, metabolomics can become a powerful tool for universal health screen.

5. Pathways and modules—power in groups

While statistical analysis of “omics” data is often penalized by false discovery rates, pathway analysis is powerful because it both brings in the context of prior knowledge and increases the statistical power while doing so [42,43]. However, the curation of pathways contains inherent human bias and is sometimes incomplete, i.e., genes of consequence are missing. In fact, pathway analysis has severe limitations when it comes to the complex data of blood transcriptomics. First, the current pathway databases are biased towards cancer, under-representing the immunology in white blood cells. Second, many pathways are based on tissues other than blood. Third, pathways poorly capture signaling cross-talks and intercellular communications. Fourth, genes in a sequential pathway may be expressed at different time, which is easily masked by heterogeneous populations of cells. Moreover, many pathways were discovered under extreme perturbations that do not reflect physiological conditions. Finally, the important context of cell types is usually missing in pathway databases.

To amend these above issues, Li et al. [46] undertook a large-scale integration of transcriptomics to define detailed molecular mechanisms in human antibody response. Using blood transcriptomics data from over 500 public studies, high-quality gene networks were reverse engineered via a mutual information approach. The resulting blood transcription modules (BTM) were validated by prior knowledge, as they recovered known protein complexes and recaptured immunological events in the literature. They also demonstrated superior sensitivity over canonical pathways. Using this new toolset of BTMs, distinct antibody response programs were identified for different types of vaccines. Examples of using BTMs as alternative to canonical pathways are shown in Fig. 2C and D, in combination with the popular GSEA software. Other efforts along this direction include a modular framework of blood genomics [75] and common axes of peripheral blood gene expression [76]. Better database curation is also under the way [49,77] (Godec et al., submitted).

The power of pathways and modules is also sought by computational metabolomics. Xia and Wishart [78] developed a metabolite set enrichment analysis, where metabolite modules were based on prior human curation. Deo et al. [79] built data-driven modules and identified a significant group of transporter reactions that escaped previous pathway curation. Li et al. [36] took the concept of metabolic pathways and networks to high-throughput metabolomics data without prior annotation. They used the collective statistical power in metabolic knowledge to resolve the ambiguity in computational prediction of metabolite identity, therefore predicting pathway and module activity in one step. This method, under the name of *mummichog*, becomes a powerful tool to accelerate metabolomics studies [80–82].

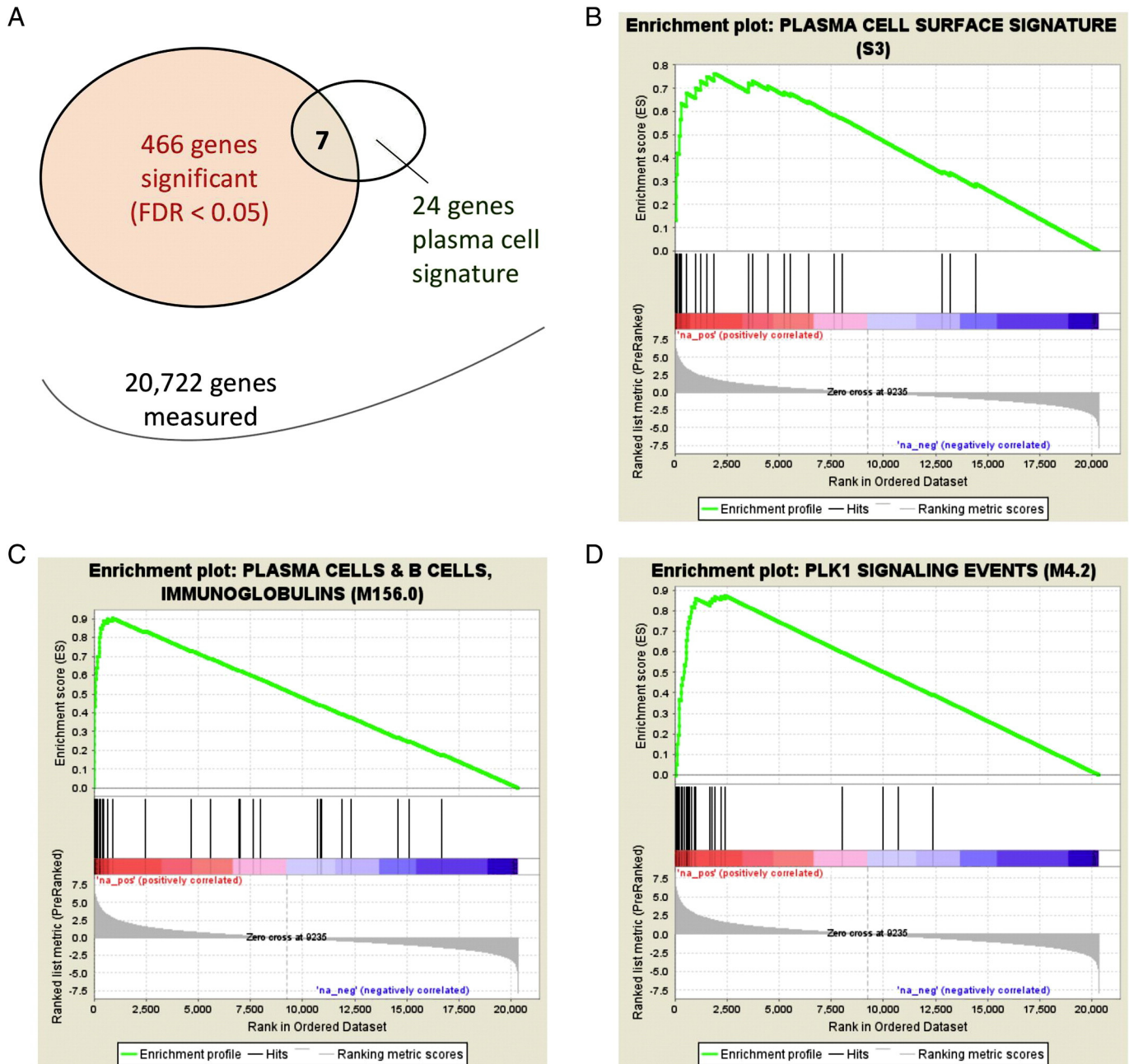


Fig. 2. Testing cell populations and gene modules in blood transcriptomics. This demonstration is based on a paired comparison between day 7 and baseline in MCIV vaccination [46]. Common statistical methods for pathway analysis are used here, while we replace conventional pathways with cell-specific signatures or custom gene modules. (A) Over-representation test. DNA microarray data are collapsed to the gene level by using the probe set of highest intensity per gene. Gene expression values are compared by paired *t*-test, and corrected for false discovery rate [50]. Among the significant genes identified here, 7 are found in a predefined signature of plasma cells. These numbers are used to construct a contingency table, and Fisher exact test returns an enrichment *p*-value < $1E^{-5}$. (B) The distribution of the same plasma cell signature genes is tested by GSEA. The bottom color bar shows the distribution of all genes, ranked by *t*-score between two time points. The vertical lines indicate the positions of the 24 genes on the ranked list, which are highly skewed for upregulation. (C) A gene module from the BTM collection [46] provides better measurement of antibody secreting cells, demonstrated on the same data. (D) Additional example of BTM module on PLK1 signaling, showing highly significant enrichment towards upregulation. The *p*-values in B, C, and D approach zero. A detailed tutorial on BTMs is available as an online supplement to Li et al. [46].

6. Integrating different data types to understand disease pathophysiology

The analysis of “omics” data is challenging and has motivated many new developments in informatics and statistics. However, each “omics” experiment only captures a static picture of dynamic and complex biology and often an averaged value of mixed signals, e.g., from many heterogeneous cells. The integration of different data types will result in a more complete understanding of disease pathophysiology and combine experimental evidences to filter out noisy signals [83–85].

Data integration can be a knowledge-driven process. For instance, enzyme proteins connect metabolites by catalyzing their conversions, and such knowledge is collected in metabolic models and databases (e.g., KEGG [86], BioCyc [87], and Reactome [88]). Guo et al. [89] recently reported that the integration of metabolomics and genomics, by matching metabolite concentration to genetic mutation on the corresponding enzymes, was successful to explain several physiological abnormalities and disease risks in relatively healthy volunteers. Genes and proteins are often conveniently organized into the annotation of genomes. In the absence of prior curation, data-driven processes become

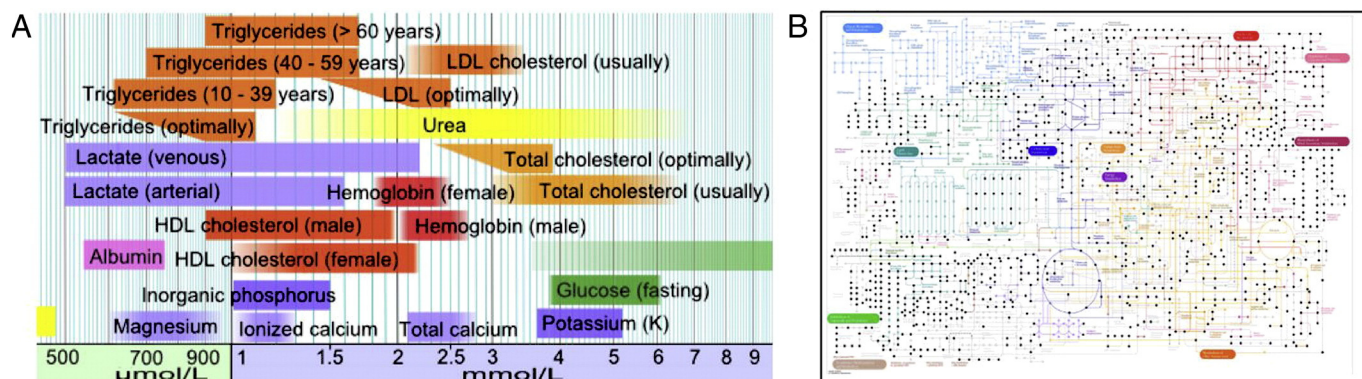


Fig. 3. Metabolomics as potential alternative to clinical blood test. (A) Partial chart of chemicals in blood test (adopted from [74]). The physiological ranges of several metabolites are shown by log scale. (B) Current coverage on KEGG pathways by LC-MS metabolomics, using data generated from our group. Each black dot is a matched metabolite. The full KEGG metabolic map can be viewed at high resolution at <http://www.genome.jp/kegg/pathway/map/map01100.html>. As metabolomics technology progresses, it can be expected to quantify over 1000 chemicals in less than 10 min. Such data will be able to support a much more detailed diagnostic chart.

necessary. For instance, transcriptomics data can be associated with genomic QTLs (quantitative trait loci) and denoted as expression QTL or “eQTL” [14,90]. Similarly, metabolomics data can support the notion of metabolomic QTL, “mQTL” [91,92].

Real-world data are often heterogeneous and require the combination of multiple methods. For example, the analysis tool for heritable and environmental network associations (ATHENA) [93] was developed to examine the associations between copy number alterations, methylation, microRNA, and gene expression with ovarian cancer survival. A neural network model was constructed for each data type separately, and the variables from the best models of each individual data set were then combined to create an integrative model using grammatical evolution neural networks (GENN) and grammatical evolution symbolic regression [94,95]. The statistical methods in ATHENA include symbolic regression, artificial neural networks, support vector machines, and GENN. These methods are selected based on a number of criteria, including fitting accuracy and robustness to non-linear interactions. Bayesian networks are also incorporated to identify conditional relationships.

Bayesian networks (BN) are a flexible and powerful method in integrating multiple “omics” data and prior information [96–101]. BNs are directed acyclic graphs in which the edges of the graph describe the conditional dependencies (given information on parent nodes) between nodes and nodes are random variables representing quantitative traits such as expression levels of genes, proteins, or metabolites. The unconnected nodes in the network represent the genes or metabolites that are conditionally independent of each other, given the parent information. Information from known interactions and pathways can be used to generate prior information of graph structure. Different weights (prior probabilities) can be given to nodes or edges reflecting researchers’ belief of the structure. Even though edges in BNs are directed, they do not represent causal relationships. However, the BN reconstruction algorithm can infer causal directions in the network by taking additional information as priors. For example, genes with *cis*-eQTLs (*cis* means locally acting on a genomic sequence) could be parent nodes of genes with coincident *trans*-eQTLs (*trans* means distally acting), but genes with *trans*-eQTLs are not allowed to be parents of genes with *cis*-eQTLs; information flows from DNA to mRNA but not in the reverse direction.

7. Concluding remarks

In the gap of common diseases and genomics, transcriptomics and metabolomics provide the important functional link and thus are key components to guide the development of personalized precision medicine. Rapid progress has been made in both areas very recently. Blood transcriptomics has now absorbed many details of human immunology.

The example of blood transcription modules [46] is a powerful tool to gauge systemic immune response from blood transcriptomics, capturing changes in both cell populations and immune pathways in general populations. Metabolomics is a fast-growing technology that captures both endogenous metabolites and environmental exposures. These data overlap with blood tests performed by current clinical methods but offer a much more powerful future alternative. The advent of these capabilities impacts many scientific and biomedical fields.

By definition, personalized medicine is an “ $n = 1$ ” problem, which however, does not mean there is less biological complexity in a single person. For that very reason, in the past few decades, the translation from animal research to clinical care has constantly seen huge disappointments. With the accumulation of detailed, information-rich data, human subjects start to contribute more to our understanding of pathobiology. It has been envisioned for some time that the combination of systems biology and epidemiology will be the prescription of personalized medicine [12]. The new developments in “blood systems biology” may be just enough to connect epidemiology, the “ $n > 1$ ” problem, to the realm of personalized medicine. That is, transcriptomics and metabolomics data from large cohorts can lead to robust models of risk factors and disease mechanisms. The future is bright also because biobank samples, even after long-term storage, can be still analyzed using newer technologies [102]. Close collaborations between computational scientists, epidemiologists and clinicians shall play a key role towards this future.

Acknowledgment

The authors appreciate stimulating discussions with Drs. Dean P. Jones and Loukia Lili-Williams. S.L. thanks research supports from the US National Institutes of Health (NIEHS P30 ES019776, NIAID 2U19 AI090023-06, NIAID HHSN272201200031C, NIA 1R01 AG038746-02, NHLBI 1P20 HL113451-01), the Department of Defense (HT9404-13-1-003), and the California Breast Cancer Research Program (21UB-8002).

References

- [1] Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12(1):56–68.
- [2] Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol* 2007;3:124.
- [3] Vogel CL, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20(3):719–26.
- [4] Kalos M, June CH. Adoptive T cell transfer for cancer immunotherapy in the era of synthetic biology. *Immunity* 2013;39(1):49–60.
- [5] Restifo NP, Dudley ME, Rosenberg SA. Adoptive immunotherapy for cancer: harnessing the T cell response. *Nat Rev Immunol* 2012;12(4):269–81.

- [6] Willis JC, Lord GM. Immune biomarkers: the promises and pitfalls of personalized medicine. *Nat Rev Immunol* 2015;15(5):323–9.
- [7] Schork NJ. Personalized medicine: time for one-person trials. *Nature* 2015; 520(7549):609–11.
- [8] Lander ES. Initial impact of the sequencing of the human genome. *Nature* 2011; 470(7333):187–97.
- [9] Manolio TA, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461(7265):747–53.
- [10] Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med* 2010; 363(4):301–4.
- [11] Clayton TA, et al. Pharmacometabonomic phenotyping and personalized drug treatment. *Nature* 2006;440(7087):1073–7.
- [12] Nicholson JK. Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* 2006;2:52.
- [13] van der Greef J, Hankemeier T, McBurney RN. Metabolomics-based systems biology and personalized medicine: moving towards $n = 1$ clinical trials? *Pharmacogenomics* 2006;7(7):1087–94.
- [14] Montgomery SB, Dermitzakis ET. From expression QTLs to personalized transcriptomics. *Nat Rev Genet* 2011;12(4):277–82.
- [15] Heng TSP. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* 2008;9(10):1091–4.
- [16] Andorf S. Towards the characterization of normal peripheral immune cells with data from ImmPort. *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics ACM*; 2014.
- [17] Brusic V. Computational resources for high-dimensional immune analysis from the Human Immunology Project Consortium. *Nat Biotechnol* 2014;32(2):146–8.
- [18] Novershtern N. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 2011;144:296–309.
- [19] Barrett T. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013;41:D991–5.
- [20] Kolesnikov N. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res* 2015;43:D1113–6.
- [21] Veenstra TD. Metabolomics: the final frontier? *Genome Med* 2012;4(4):40.
- [22] Jones DP, Park Y, Ziegler TR. Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr* 2012;32:183–202.
- [23] Hopfgartner G, Tonoli D, Varesio E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Anal Bioanal Chem* 2012;402(8):2587–96.
- [24] Kaufmann A, et al. Comprehensive comparison of liquid chromatography selectivity as provided by two types of liquid chromatography detectors (high resolution mass spectrometry and tandem mass spectrometry): “Where is the crossover point?”. *Anal Chim Acta* 2010;673(1):60–72.
- [25] Smith CA. XCMS: processing mass spectrometry data for metabolite profiling using non-linear peak alignment, matching and identification. *Anal Chem* 2006;78.
- [26] Sturm M. OpenMS—an open-source software framework for mass spectrometry. *BMC Bioinf* 2008;9:163.
- [27] Yu T, et al. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* 2009;25(15):1930–6.
- [28] Uppal K, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinf* 2013;14:15.
- [29] Wishart DS, et al. HMDB: the human metabolome database. *Nucleic Acids Res* 2007;35(Database issue):D521–6.
- [30] Wishart DS, et al. HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res* 2013;41(Database issue):D801–7.
- [31] Smith CA. METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005; 27(6):747–51.
- [32] Austin CP. NIH molecular libraries initiative. *Science* 2004;306(5699):1138–9.
- [33] Degtyarenc K. ChEBI: an open bioinformatics and cheminformatics resource. *Curr Protoc Bioinformatics* 2009;14:14.9.
- [34] Gowda H. Interactive XCMS online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal Chem* 2014;86(14):6931–9.
- [35] Xia J. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res* 2015;43(W1):W251–7.
- [36] Li S, et al. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* 2013;9(7), e1003123.
- [37] Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;14(8):1847–50.
- [38] Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. *Toxicol Sci* 2014;137(1):1–2.
- [39] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 1995; 289–300.
- [40] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2001;1:165–88.
- [41] Allison DB. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006;6(1):55–65.
- [42] Li S. Systems biological approaches to measure and understand vaccine immunity in humans. *Semin Immunol* 2013;25(3):209–18.
- [43] Khatri P. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 2012;8(2).
- [44] Abbas AR. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One* 2009;4(7), e6098.
- [45] Shen-Orr SS. Cell type-specific gene expression differences in complex tissues. *Nat Methods* 2010;7(4):287–9.
- [46] Li S, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 2014;15(2):195–204.
- [47] Bolen CR. Cell subset prediction for blood genomic studies. *BMC Bioinf* 2011; 12(258).
- [48] Nakaya HI. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* 2011;12(8):786–95.
- [49] Subramanian A. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102: 15545–50.
- [50] Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100(16):9440–5.
- [51] Spratlin JL. Clinical applications of metabolomics in oncology: a review. *Clin Cancer Res* 2009;15.
- [52] Armitage EG. Metabolomics in cancer biomarker discovery: current trends and future perspectives. *J Pharm Biomed Anal* 2014;87.
- [53] Folger O. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* 2011;7.
- [54] Chiarugi A, et al. The NAD metabolome—a key determinant of cancer cell biology. *Nat Rev Cancer* 2012;12:11:741–54.
- [55] Halama A, et al. Identification of biomarkers for apoptosis in cancer cell lines using metabolomics: tools for individualized medicine. *J Intern Med* 2013;274:5:425–39.
- [56] Zhao J, et al. Novel metabolic markers for the risk of diabetes development in American Indians. *Diabetes Care* 2015;38:2:220–7.
- [57] Xia J. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* 2013;9.
- [58] Wheelock CE. Application of ‘omics technologies to biomarker discovery in inflammatory lung disease. *Eur Respir J* 2013;42.
- [59] Smolinska A. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Anal Chim Acta* 2012;750.
- [60] Mamas M. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch Toxicol* 2011;85.
- [61] Wang TJ. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17.
- [62] Suhre K, et al. Metabolic footprint of diabetes: a multipplatform metabolomics study in an epidemiological setting. *PLoS One* 2010;5(11), e13953.
- [63] Kim OY. Metabolomic profiling as a useful tool for diagnosis and treatment of chronic disease: focus on obesity, diabetes and cardiovascular diseases. *Expert Rev Cardiovasc Ther* 2013;11.
- [64] Trushina E, Mielke MM. Recent advances in the application of metabolomics to Alzheimer’s Disease. *Biochim Biophys Acta* 2014;1842:8:1232–9.
- [65] Cano A, Alonso C. Deciphering non-alcoholic fatty liver disease through metabolomics. *Biochem Soc Trans* 2014;42(5):1447–52.
- [66] Roberts LD, Koulman A, Griffin JL. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol* 2014;2(1):65–75.
- [67] Osborn MP, et al. Metabolome-wide association study of neovascular age-related macular degeneration. *PLoS One* 2013;8(8), e72737.
- [68] Fitzpatrick AM, et al. Children with severe asthma have unique oxidative stress-associated metabolomic profiles. *J Allergy Clin Immunol* 2014;133(1):258–61, e1–8.
- [69] Roede JR, et al. Serum metabolomics of slow vs. rapid motor progression Parkinson’s disease: a pilot study. *PLoS One* 2013;8(10), e77629.
- [70] Kalhan SC, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism* 2011;60(3):404–13.
- [71] Weiner 3rd J, et al. Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS One* 2012;7(7), e40221.
- [72] Park YH, et al. High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. *Toxicology* 2012;295(1–3):47–55.
- [73] Miller MJ, et al. Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. *J Inher Metab Dis* 2015;38(6):1029–39.
- [74] Haggstrom M. Medical gallery of Mikael Häggström 2014. *Wikiversity J Med* 2014; 1(2).
- [75] Chaussabel D, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008;29(1):150–64.
- [76] Preiner M, et al. Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet* 2013;9(3), e1003362.
- [77] Lynn DJ. InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol* 2008;4:218.
- [78] Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* 2010;38(Suppl. 2):W71–7.
- [79] Deo RC. Interpreting metabolomic profiles using unbiased pathway models. *PLoS Comput Biol* 2010;6(2).
- [80] Hoffman JM, et al. Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, *Drosophila melanogaster*. *Aging Cell* 2014; 13(4):596–604.
- [81] Xu X, et al. Autophagy is essential for effector CD8(+) T cell survival and memory formation. *Nat Immunol* 2014;15(12):1152–61.
- [82] Johnson C. Bioinformatics: the next frontier of metabolomics. *Anal Chem* 2014; 87(1):147–56.
- [83] Joyce AR, Pálsson BØ. The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol* 2006;7(3):198–210.
- [84] Ritchie MD, et al. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* 2015;16(2):85–97.
- [85] Topol EJ. Individualized medicine from prewomb to tomb. *Cell* 2014;157(1): 241–53.

- [86] Kanehisa M, et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014;42(Database issue):D199–205.
- [87] Karp PD. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 2005;33.
- [88] Matthews L. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* 2009;37.
- [89] Guo L, et al. Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *Proc Natl Acad Sci U S A* 2015;112(35):E4901–10.
- [90] Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet* 2006;7(11):862–72.
- [91] Shin SY, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;46(6):543–50.
- [92] Dumas M-E, Gauguier D. Mapping metabolomic quantitative trait loci (mQTL): a link between metabolome-wide association studies and systems biology. *Genetics Meets Metabolomics* Springer; 2012. p. 233–54.
- [93] Holzinger E. ATHENA: the analysis tool for heritable and environmental network associations. *Bioinformatics* 2014;30(5):698–705.
- [94] Turner SD. ATHENA: a knowledge-based hybrid backpropagation-grammatical evolution neural network algorithm for discovering epistasis among quantitative trait loci. *Biodata Min* 2010;3(1).
- [95] Holzinger ER. ATHENA: a tool for meta-dimensional analysis applied to genotypes and gene expression data to predict HDL cholesterol levels. *Pac Symp Biocomput* 2013:385–96.
- [96] Friedman N. Using Bayesian networks to analyze expression data. *J Comput Biol* 2000;7:601–20.
- [97] Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 2005;308(5721):523–9.
- [98] Yeung KY. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* 2005;21(10):2394–2402.
- [99] Xing H. Causal modeling using network ensemble simulations of genetic and gene expression data predicts genes involved in rheumatoid arthritis. *PLoS Comput Biol* 2011;7.
- [100] Zhu J. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol* 2012;10(4).
- [101] Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature* 2009;461(7261):218–23.
- [102] Hebel DG, et al. Performance in omics analyses of blood samples in long-term storage: opportunities for the exploitation of existing biobanks in environmental health research. *Environ Health Perspect* 2013;121(4):480–7.